International Journal of Research and Applied Innovations (IJRAI)

| ISSN: 2455-1864 | www.ijrai.com | editor@ijrai.com | A Bimonthly, Scholarly and Peer-Reviewed Journal |

||Volume 7, Issue 2, March–April 2024||

DOI:10.15662/IJRAI.2024.0702001

AI-Augmented ITSM: Autonomous Incident Triage

Nilesh Vijay Patel

Sat Kabir Institute of Technology and Management, Haryana, India

ABSTRACT: Modern IT environments face increasing complexity, with high volumes of incidents making manual triage inefficient and error-prone. AI-augmented ITSM (IT Service Management) systems offer promise by automating classification, prioritization, routing, and resolution of incidents—ultimately reducing mean time to resolution (MTTR) and enhancing service reliability. This paper examines prior-2022 advancements in autonomous incident triage, focusing on AI methods that support decision-making within ITSM.

We analyze **DeepTriage** (Microsoft Azure), an ensemble of gradient-boosted trees, clustering, and deep networks deployed in cloud incident categorization, achieving high F1 scores (82.9%) in production environments across thousands of teams arXiv. **SoftNER**, used at Microsoft, extracts structured knowledge (entities like system components and error codes) from incident reports via BiLSTM-CRF, improving downstream triage accuracy arXiv. **DeCaf**, another Microsoft system, automates diagnosis and triaging of KPI-based performance regressions using machine learning and pattern mining, effectively surfacing root causes from log data arXiv. Additionally, research has demonstrated that multi-modal analysis—incorporating images along with text—enhances routing and resolution outcomes in IT support arXiv.

We synthesize these contributions into a unified methodology: leveraging multi-modal input, entity extraction, predictive routing, and root-cause diagnosis in an autonomy-capable ITSM pipeline.

Advantages include higher triage speed, consistency, and scalable performance under high incident loads. Disadvantages include model trust and explainability challenges, data quality dependencies, integration hurdles, and monitoring needs.

The study shows that while fully autonomous incident handling remains aspirational, AI-augmented triage systems have already delivered significant operational improvements. Future directions involve enhancing explainability, expanding multimodal understanding, integrating real-time monitoring (AIOps), and supporting closed-loop automation.

KEYWORDS: AI-Augmented ITSM. Autonomous Incident Triage, Incident Classification, DeepTriage, SoftNER, DeCaf, Multi-modal Analysis, Machine Learning in ITSM, Root-Cause Analysis, Knowledge Extraction

I. INTRODUCTION

Incident triage is a cornerstone of IT Service Management (ITSM), determining the right actions or team assignments to resolve system issues efficiently. As IT infrastructures expand—cloud services, IoT, microservices—the volume and complexity of incidents surge. Traditional manual triage struggles under this load, leading to long response times and inconsistent resolution quality.

AI augments ITSM by automating triage through learning from historical data. The goal is to mimic expert decision-making—classifying, prioritizing, and routing incidents—with greater speed and consistency. Prior to 2022, several successful systems emerged:

- **DeepTriage**, deployed in production at Microsoft Azure since 2017, leverages ensemble learning to recommend responsible teams with high F1 scores arXiv.
- **SoftNER**, also at Microsoft, applies NLP to extract structured knowledge from service incidents, facilitating effective routing and triage improvements arXiv.

| ISSN: 2455-1864 | www.ijrai.com | editor@ijrai.com | A Bimonthly, Scholarly and Peer-Reviewed Journal |

||Volume 7, Issue 2, March–April 2024||

DOI:10.15662/IJRAI.2024.0702001

- **DeCaf** uses pattern mining and ML to diagnose performance regressions from logs, enabling automated identification of root causes arXiv.
- Multi-modal analysis (text + images) has been shown to enrich incident understanding and improve routing accuracy arXiv.

These systems exemplify how AI can shoulder significant triage responsibilities—rendering IT operations more proactive and adaptive. This paper evaluates these methods, identifies strengths and weaknesses, and proposes a unified AI-augmented ITSM architecture.

II. LITERATURE REVIEW

DeepTriage (2020): A production-grade AI triage system at Microsoft Azure combining gradient-boosted decision trees, clustering, and deep neural networks. It handles imbalanced incident distributions, variable input formats, and meets production scalability. Achieved an overall F1 score of 82.9%, ranging from 76.3% to 91.3% on high-impact incidents arXiv.

SoftNER (2020): An unsupervised framework for extracting structured knowledge (e.g., entities) from incident descriptions using a multi-task BiLSTM-CRF model, trained via bootstrapping on key-value structures. Deployed at Microsoft, yielding a 0.96 precision and improving downstream triage models arXiv.

DeCaf (2019): A system for diagnosing performance KPI regressions in cloud services. Applies ML and pattern mining over logs to both diagnose and triage issues in Microsoft services. Demonstrated capability to scale and integrate into DevOps environments arXiv.

Multi-Modal Incident Analysis (2019): Incorporating images (e.g., screenshots) along with text into triage systems leads to improved routing and resolution. Evaluated on ~25,000 real IT tickets, multi-modal input yielded significant accuracy improvements over text-only systems arXiv.

These works illustrate a progression—from structured prediction (DeepTriage) to enriched semantic understanding (SoftNER) to log-based diagnosis (DeCaf), and to multimodal comprehension. They collectively inform a comprehensive AI-augmented triage pipeline.

III. RESEARCH METHODOLOGY

To design an autonomous incident triage system, we propose the following multi-stage methodology:

Multi-modal Data Ingestion
Collect incident reports that may include text descriptions, log excerpts, images/screenshots, and performance metrics.

Knowledge Extraction (SoftNER-style)
Apply NLP models (e.g., BiLSTM-CRF) to extract key entities—incident type, affected systems, error codes—to structure the input.

Multi-Modal Feature Fusion
Combine textual, visual, and extracted entity features into unified embeddings for richer context.

Incident Classification & Routing (DeepTriage-style)
Employ ensemble models (gradient-boosted trees, neural networks, clustering) to classify incidents and predict responsible teams or routing paths.

Root-Cause Diagnosis (DeCaf-style)
Integrate log-pattern mining and KPI correlations to suggest probable root causes, leveraging historical trend analysis.

Feedback Loop & Model Updating
Capture post-resolution data to retrain models, improving accuracy and adapting to evolving patterns (a DevOps feedback principle).

International Journal of Research and Applied Innovations (IJRAI)

| ISSN: 2455-1864 | www.ijrai.com | editor@ijrai.com | A Bimonthly, Scholarly and Peer-Reviewed Journal |

||Volume 7, Issue 2, March–April 2024||

DOI:10.15662/IJRAI.2024.0702001

Scalability and Integration
Ensure this pipeline operates in production-grade environments—low latency, resilience, and compatibility with ITSM tools.

Human Oversight & Trust
Provide explainability ("why this triage decision?") and enable manual review for high-impact incidents.
This layered methodology fuses best practices to yield autonomous, reliable, and explainable triage.

Advantages

- **Faster Triage and Reduced MTTR:** Quickly classify and route incidents to the right teams.
- **Consistency and Scalability:** Eliminates variability in human judgment, scales with incident volume.
- **Improved Diagnosis:** Log-based insights and entity extraction enhance root-cause discovery.
- **Rich Context Understanding:** Multimodal inputs capture broader incident context.
- **Continuous Improvement:** Feedback loop enables adaptive learning and evolving performance.

Disadvantages

- **Data Quality Dependency:** Requires high-quality labeled historical incidents; noisy or sparse data degrades performance.
- **Trust and Interpretability:** Black-box models reduce transparency; manual overrides remain necessary.
- **Integration Overhead:** Complex to integrate with legacy ITSM tools and workflows.
- **Resource and Maintenance Costs:** Training and updating models require effort and expertise.
- **Coverage Gaps:** Rare or novel incident types may be misclassified without fallback human handling.

IV. RESULTS AND DISCUSSION

- **DeepTriage** improved team assignment accuracy within Azure operations, handling thousands of incidents daily while maintaining 80–90% F1 performance arXiv.
- **SoftNER** enabled structured knowledge extraction from raw text, bolstering triage model performance in production context arXiv.
- **DeCaf** successfully diagnosed both known and unknown performance issues, demonstrating efficacy in real-world cloud platforms arXiv.
- **Multi-modal analysis** significantly improved resolution and routing metrics over text-based systems in a real incident ticket corpus arXiv.
- These outcomes validate that combining multi-modal understanding, structured knowledge, and diagnostic reasoning can materially enhance autonomous triage.

V. CONCLUSION

AI-augmented incident triage offers substantial gains in speed, consistency, and diagnostic power. Systems like DeepTriage, SoftNER, and DeCaf illustrate the feasibility and benefits of stacking classification, knowledge extraction, and log analytics into an autonomous pipeline. Yet, challenges in trust, integration, and data readiness persist. A balanced design that retains human oversight and prioritizes explainability remains essential.

VI. FUTURE WORK

1. **Explainable AI Models:** Develop interpretable triage decisions (e.g., attention tracking, rationale summaries).
2. **Expanded Multimodal Inputs:** Integrate metrics like performance telemetry, code changes, and user feedback.
3. **Real-time Incident Prediction:** Shift from reactive triage to predictive incident prevention.
4. **Adaptive Learning in Production:** Online learning frameworks responding to new incident patterns.
5. **Cross-Domain Transfer Learning:** Apply models across different IT environments or incident types.

| ISSN: 2455-1864 | www.ijrai.com | editor@ijrai.com | A Bimonthly, Scholarly and Peer-Reviewed Journal |

||Volume 7, Issue 2, March–April 2024||

DOI:10.15662/IJRAI.2024.0702001

REFERENCES

1. Pham, P., Jain, V., Dauterman, L., Ormont, J., & Jain, N. (2020). *DeepTriage: Automated Transfer Assistance for Incidents in Cloud Services*. arXiv (turn0academia13).
2. Shetty, M., Bansal, C., Kumar, S., Rao, N., Nagappan, N., & Zimmermann, T. (2020). *Neural Knowledge Extraction From Cloud Service Incidents*. arXiv (turn0academia15).
3. Bansal, C., Renganathan, S., Asudani, A., Midy, O., & Janakiraman, M. (2019). *DeCaf: Diagnosing and Triaging Performance Issues in Large-Scale Cloud Services*. arXiv (turn0academia14).
4. Mandal, A., Agarwal, S., Malhotra, N., Sridhara, G., Ray, A., & Swarup, D. (2019). *Improving IT Support by Enhancing Incident Management Process with Multi-modal Analysis*. arXiv (turn0academia12).